
Status on the Mapping of Metadata Standards: ISO/IEC 11179, SDMX, and Others

7 April, 2008

Arofan Gregory

Contents

I. Introduction	2
II. The Mapping Problem.....	3
III. Current Status of Mappings	5
A. SDMX-ISO/IEC 11179	5
B. DDI - ISO/IEC 11179	7
C. SDMX - DDI	9
D. XBRL-SDMX.....	9
E. XBRL-ISO/IEC 11179	10
IV. On-Going Work.....	10
V. Summary	11

I. Introduction

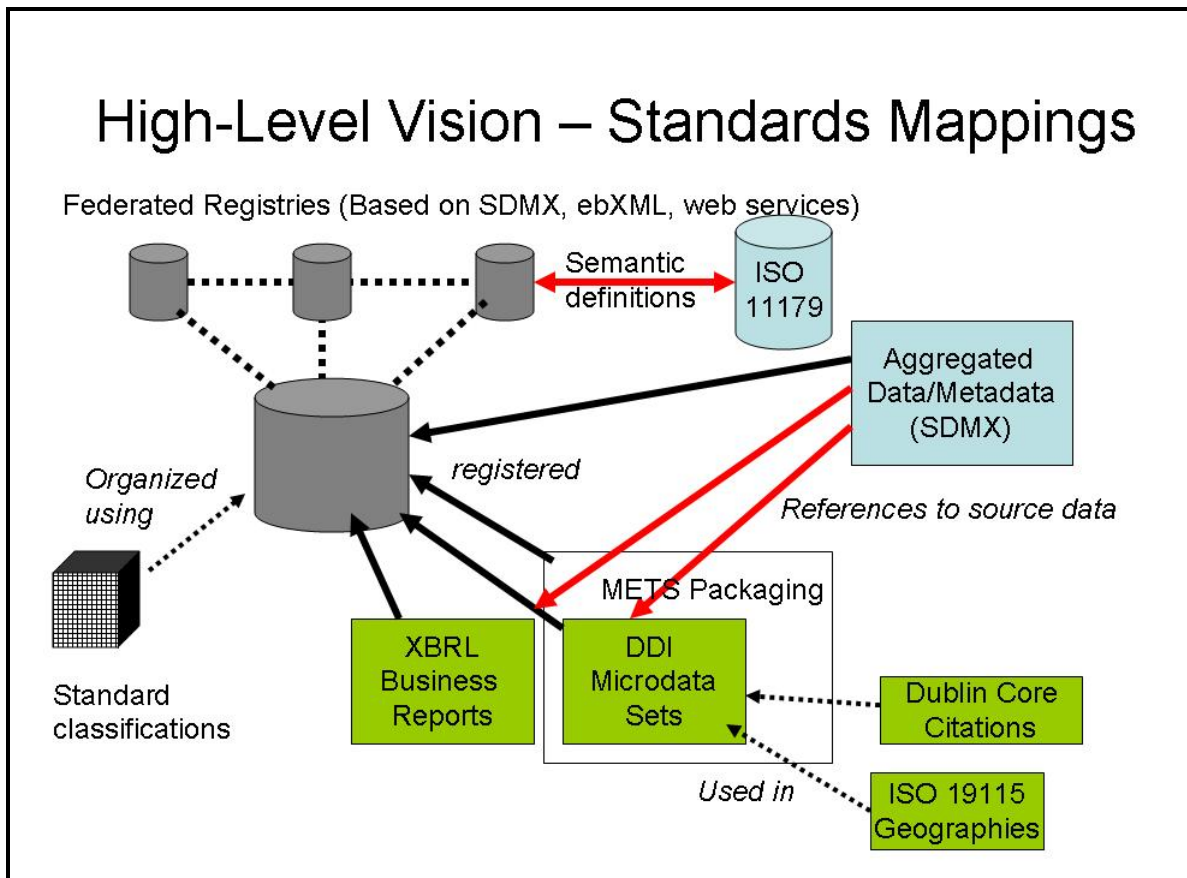
At the METIS meeting in 2006, several papers were presented concerning some of the standards which are important for statistics and related metadata, and many discussions were held by attendees. Standards of note included ISO/IEC 11179, the Statistical Data and Metadata Exchange standards (SDMX), the Data Documentation Initiative (DDI), ISO 19115 (geographical metadata), Dublin Core, and the eXtensible Business Reporting Language (XBRL). It is the purpose of this paper to give an update of the status of attempts to map these standards to one another, with a focus on SDMX and ISO/IEC 11179, but also providing some information regarding the other mentioned standards.

There has been on-going discussion and work regarding many aspects of this mapping, and some important progress has been made. When trying to map metadata standards, there are almost always several different approaches which are equally viable. As with many problems in the application of technology, one of the key questions which must be answered is "What are we trying to achieve?". By defining concrete use cases, a basis for mapping standards against one another is formed, and the work is provided with criteria for assessment. If the intended use cases are supported, then it is possible to determine that the standards have been mapped "correctly". Much of the work in this past year has focused on concrete use cases which provide this kind of underpinning to the overall work.

It is not insignificant that SDMX has been a focal point for this effort: because the SDMX version 2.0 technical specifications will be submitted to ISO in 2008, through TC-154, it has been necessary to establish the relationship between SDMX and other standards of interest to that committee (including ISO/IEC 11179, some of the ISO 15000 ebXML standards, XBRL, and DDI). Further, DDI has just submitted their version 3.0 specification for a final vote, with the expectation that this major increase in the scope of that specification will be approved in May of 2008. These (and other) developments have highlighted the need for concrete mappings between standards, as we are reaching a point where a coherent set of coordinated standards is becoming a reality for data collection and statistical reporting and dissemination.

II. The Mapping Problem

The first challenge facing the work of standards mapping includes defining relevant use cases. If we can understand what is a realistic goal for using data and metadata expressed in one standard in relation to another, then we can begin to understand how one approach is to be preferred to another. Although we cannot claim definitively that all the useful cases have been formulated, we have defined a useful set of cases which serve as an underpinning for the on-going work. The overall picture is presented in the diagram below.



In 2006, the initial proposal was to try to map all of the standards which were of concern (ISO/IEC 11179, SDMX, DDI, ISO 19115, XBRL, Dublin Core) against ISO/IEC 11179, so that this standard served as a pivotal model for all of the others. This approach was used because of the nature of ISO/IEC 11179 - it provides a set of attributes for defining the meanings and representations of terms, concepts, and data elements, and thus naturally fits into this role when mapping standards.

This approach has in some cases proven to be very useful, and in other cases has proven to only partially fill the need as established by our use cases. It should be understood that mapping all of the mentioned standards against one another is a huge effort - it is important that the work be focused, so that a useful result can be achieved. Relying on ISO/IEC 11179 as a pivotal standard is still seen as the correct approach, but it has in some cases been supplemented by direct mappings between two of the other standards.

When we consider the diagram, we can see the overall vision: we have two basic forms of micro-data which are important: on the business side, we have XBRL business reports, and in the area of survey data and data sourced from registers, we have the micro-data files documented in DDI. This latter standard includes native, embedded Dublin Core citations, and provides for links to ISO 19115-compliant shape files for expressing geographical links.

Such micro-data forms the input to aggregate data. Given a registry mechanism such as that provided by the ISO 15000 ebXML and SDMX standards, we can visualize a network of federated registries which allow for links between the source micro-data (XBRL, DDI, etc.) and the aggregates which are expressed according to the SDMX model. ISO/IEC registries can provide a powerful resource within this network, by allowing for a standard definition of concepts and representations, and the means to manage these resources throughout their lifecycle. Some additional standards form part of this picture: METS (a digital archiving standard), web-services technology standards, and several of the standards involved in classification management. These are not the subject of the current paper, but are potentially significant in implementations, and deserve at least a passing mention.

Within this overall vision are several important mappings:

(1) DDI-SDMX: This is important because if we are to use an SDMX registry mechanism as the basis of our coordination throughout the network, we must be able to cross-walk between DDI metadata and native SDMX objects.

(2) XBRL-SDMX: This is a similar case to the DDI case - we must be able to cross-walk between XBRL inputs and known SDMX objects in order to use the registry mechanism of SDMX for these micro-data sources.

(3) SDMX-ISO/IEC 11179: This is important also, as the integration of an ISO/IEC 11179 registry into our network is not possible without having a known

correspondence between the SDMX objects and their meanings and representations in ISO/IEC 11179.

(4) DDI-ISO/IEC 11179: Because we have DDI objects registered on the network, it is necessary to understand how they integrate with ISO/IEC 11179 as well.

(5) XBRL-ISO/IEC 11179: Because XBRL data and metadata exist as registered objects within our network, these also must be definable according to the ISO/IEC 11179 model.

This picture highlights this set of mappings as initial starting points. It is important to note that some of the standards which are considered important - Dublin Core and ISO 19115 - already have partial mappings implicit in the fact that they are embedded within DDI. Although this is not sufficient for a complete picture, both standards can be described also as reference metadata sets in SDMX terminology. Thus, this area is seen as one which will not be difficult to manage.

III. Current Status of Mappings

A. SDMX-ISO/IEC 11179

The general use case for this mapping is that any concept or representation found in an SDMX object (a data set, a structural metadata set, or a reference metadata set) should be describable in ISO/IEC 11179 terms, so that an application could leverage an ISO/IEC 11179 registry in helping to manage the content. It is easy to visualize applications which use this functionality: if I am designing a data structure definition in SDMX, it would be very useful to have access to existing concepts and code lists defined within an ISO/IEC 11179 registry, so that I could use the full set of information in understanding and selecting concepts and code lists. Users of the resulting data would also benefit from having access to the ISO/IEC 11179 metadata associated with the SDMX concepts and codes.

It is notable that the design of SDMX intentionally does not duplicate the detailed nature of ISO/IEC 11179 in terms of its property set. The two standards are well-suited for working together in this type of application. the diagram below shows how the less-detailed use of concepts and representations within SDMX maps to the constructs within ISO/IEC 11179.

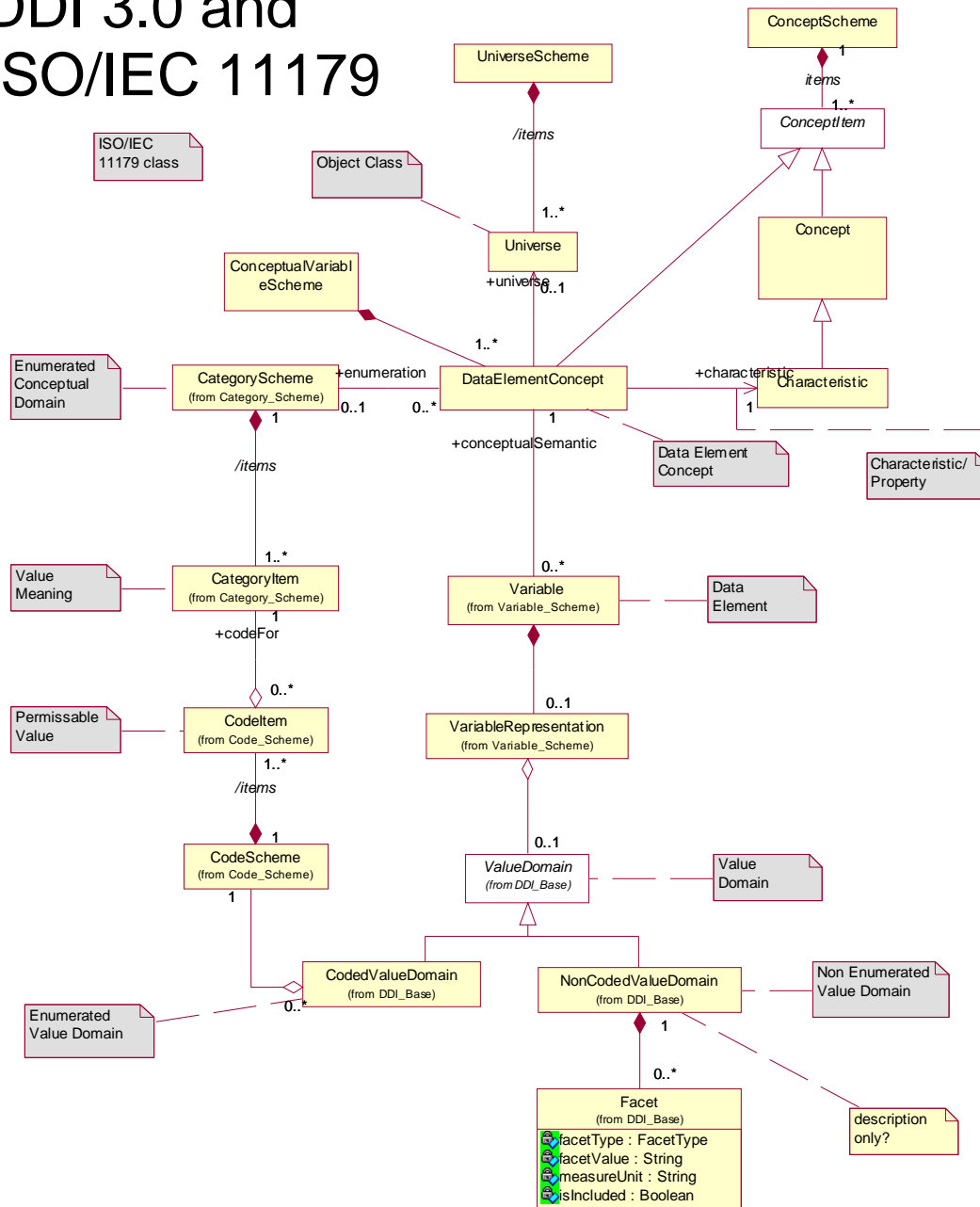
To summarize, the critical mapping between SDMX and ISO/IEC 11179 has been fundamentally established. Although a formal expression of this mapping does not yet exist in any but draft form, the really difficult work has already been done. We anticipate that a formal expression of this mapping will be available for comment sometime in 2008. This is great news, because this has been one of the more difficult of the mappings to understand and realize.

B. DDI - ISO/IEC 11179

DDI has a privileged relationship with ISO/IEC 11179 because the ISO/IEC 11179 model was used as the basis for many of the corresponding metadata structures within DDI. At the 2007 IASSIST conference in Montreal, much work was done to make sure that ISO/IEC 11179 concepts and representations could be aligned with the metadata payload of DDI 3.0. The UML diagram below shows how this works.

It is not the intention of this paper to go into the full detail of this mapping, but it is useful to describe the overall principle. DDI provides an XML format for metadata regarding micro-data sets. It is possible, based on the provided mapping, to express the relevant DDI constructs – code lists, categories, concepts, and universes - as the corresponding constructs in the ISO/IEC 11179 model. Thus, DDI 3.0 can be employed as a syntax implementation of ISO/IEC 11179. There is no requirement that DDI 3.0 be used in this way, but the integration of the two metadata standards is essentially complete. This technique has been described in the documentation for the use of the DDI version 3.0 XML syntax, and a more complete description of the mapping is currently being undertaken.

DDI 3.0 and ISO/IEC 11179



C. SDMX - DDI

This is one of the mappings which is required by the defined use cases, but which does not employ ISO/IEC as a pivotal format. Because both SDMX and DDI have constructs which are described in ISO/IEC 11179, the pivotal function of that standard covers much of the needed territory: code lists and other representations, categories, and concepts are all mappable using ISO/IEC 11179. However, DDI and SDMX also cover some areas which ISO/IEC 11179 has no coverage of, notably the use of representations and concepts in the higher-level structuring of data sets and metadata sets. Prototyping has shown that there are several areas where SDMX and DDI can usefully be mapped: both are capable of describing multi-dimensional matrixes for data, and expressing not only the data structures in an XML syntax but also capturing the data itself. It should be mentioned that DDI 3.0 was developed with the SDMX model in mind, so that a cross-walk will be possible in this regard.

Recent work has also shown that other aspects of DDI can be expressed usefully in SDMX: notably, the DDI lifecycle model can be expressed as an SDMX "Process", with each phase of the lifecycle seen as an SDMX "Process Step". This has important implications, in that any DDI metadata which applied to aggregate data expressed in SDMX could be persisted in the form of an SDMX reference metadata report associated with a step in this lifecycle. Further, a mapping is now being tested which would allow for DDI metadata to be automatically mapped into SDMX data structures, so that micro-data sets could be expressed as SDMX data sets. Although micro-data has limited dimensionality, this cross-walk allows for any SDMX tools which display data to also display micro-data. This work is not yet complete, but an initial example has been created, and some experiments are now being undertaken. If this mapping can be established, it means that native SDMX tools for working with data could also leverage DDI metadata and micro-data, based on transformations. Given our use cases, this would be very powerful in allowing the users of SDMX to leverage DDI sources.

D. XBRL-SDMX

There has been an agreed project between the SDMX Sponsors and XBRL International to work on a formal, standard mapping between these two standards. A description of this work is beyond the scope of this paper, but much progress has been made in this area, and we anticipate that there will be further meetings in 2008 to examine the informal progress which has been made by members of these two groups. An initial draft has been created, which hopefully

will become the subject of a more formal meeting in the next few months. Progress in this area can best be monitored by checking the SDMX.org website. Prospects at this time seem very promising, however.

E. XBRL-ISO/IEC 11179

This is one area where no work is yet being undertaken. It is easy to assume that an XBRL taxonomy could be expressed in terms of the ISO/IEC 11179 model, but no actual examples or prototypes yet exist in this area. Hopefully, this mapping will be undertaken in the coming year.

IV. On-Going Work

Members of several organizations have been active in the work being described. Experts from the SDMX community, from METIS, from XBRL International, from the DDI Alliance, the Open Data Foundation, ISO TC-154, and from various other ISO committees attending the Open Forum for Metadata Registries have participated in informal discussions and work at different stages in the past two years, as well as individuals from several national statistical offices. The degree of interest in this work has been very heartening, and we feel that solid progress is being made.

Based on the status of work as described above, the following work products should be forthcoming, as drafts for review, in the course of 2008:

- A documented mapping between ISO/IEC 11179 and SDMX
- A documented mapping between ISO/IEC 11179 and DDI
- A documented mapping between SDMX and DDI
- A documented mapping between XBRL and SDMX

This covers 4 of the 5 identified mappings needed to support our use cases, so progress here is seen as significant. In most cases, the work is being undertaken in external (and often informal) groups, with the intention that end results can be submitted to METIS for inclusion in the metadata framework. Exceptions to this are the XBRL-SDMX mapping, which will presumably be jointly published by the SDMX Sponsors and XBRL International, and which could be referenced by METIS in its own publications.

It is exciting to see that not only are the mappings themselves emerging, but that software tools to realize some of these cross-walks are also being prototyped. Hopefully, 2008 will see more implementation of the mappings, both to prove that

they are real and useful, and also to possibly provide open-source tools for use in implementations.

V. Summary

Since the work was initiated at METIS 2006, much progress has been made, and 2008 should see the circulation of several drafts for comment. This work is sometimes difficult and resource-intensive, and is often not seen as mission-critical by individual organizations, so the degree of progress which has been made is especially heartening. What has been accomplished forms a solid basis for further work, but it is clear that much work remains to be done: an XBRL-ISO/IEC 11179 mapping has not yet been undertaken, and several additional mappings - particularly ISO 19115 and Dublin Core - will also be needed.

Perhaps the most positive aspect of this effort has been the support provided by the various standards groups themselves. It is clear that many of the standards committees and organizations such as the SDMX Sponsors, the DDI Alliance, members of many ISO committees, and XBRL International are all aware of the need to provide mappings between their specifications. The METIS community can clearly benefit from this awareness, and help in the coordination and publication of the results of this work.